AI & ML

MODULE 2

SESSION 6

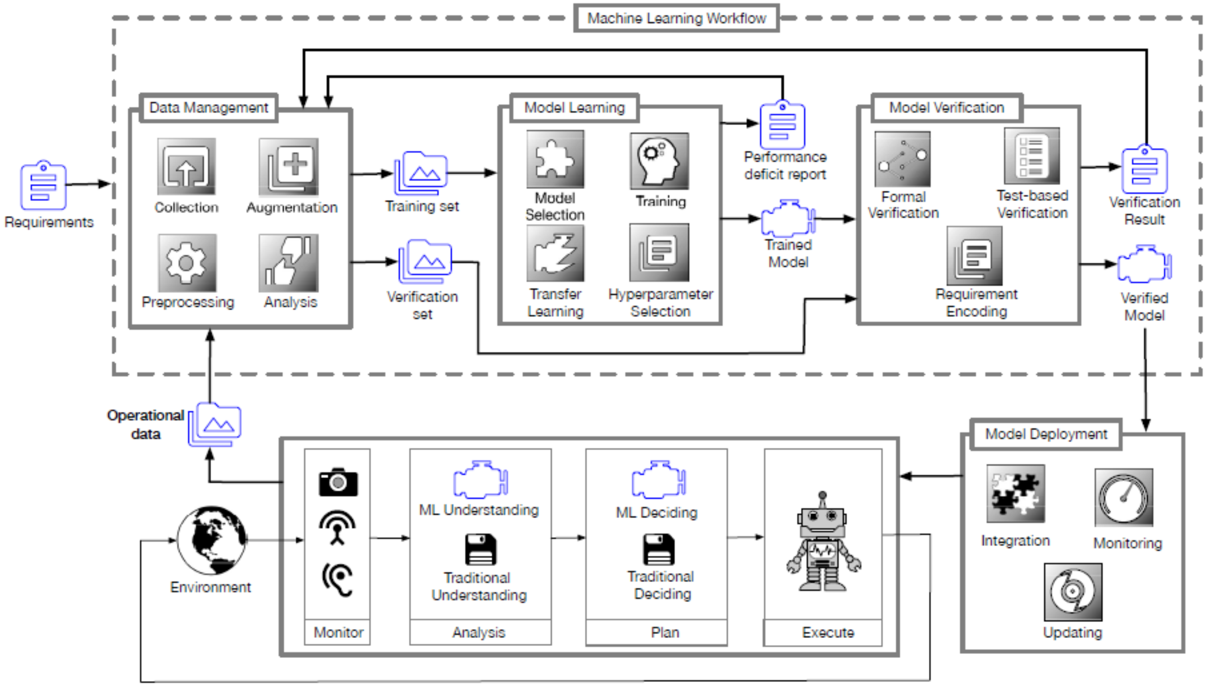
Session Outline

- ML Lifecycle
- Types of ML
- Terminologies
- Exploratory Data Analysis

ML lifecycle

Machine Learning Lifecycle

Planning → Data Management → Model Learning → Model Verification → Model Deployment → Monitoring and Maintenance



Planning

Planning

- assessing the scope, success metric, and feasibility of the ML application.
- understand the cost-benefit analysis
- define clear and measurable success metrics for business, machine learning models, and economic.
- Consists of two steps
  - Understanding of the problem
  - Framing an ML problem

Understanding the problem

1. State the goal for the product being developed.

Application	Goal
Weather App	Calculate precipitation in six-hour increments for a geographic region
Mail App	Detect spam
Banking App	Identify Fraudulent Transaction
Video App	Recommend useful videos

2. Determine whether the goal is best solved using ML.
3. Verify whether the data required to train a model is available.

Framing an ML Problem

1. Define the ideal outcome and the model's goal.

App	Ideal outcome	Model's goal
Weather app	Calculate precipitation in six hour increments for a geographic region.	Predict six-hour precipitation amounts for specific geographic regions.
Video app	Recommend useful videos.	Predict whether a user will click on a video.
Mail app	Detect spam.	Warn the user if the email appears to be spam.
Banking app	Identify fraudulent transactions.	Predict if a transaction was made by the card holder.

2. Identify the model's output.
3. Define success metrics.

Why is problem framing important?

A. Problem framing helps diagnose problems with existing ML models and uncovers issues with data.

B. Problem framing ensures that an ML approach is a good solution to the problem before beginning to work with data and train a model.

A fashion firm wants to sell more clothes. Someone suggests using ML to determine which clothes the firm should manufacture. They think they can train a model to determine which type of clothes are in fashion. After they train the model, they want to apply it to their catalog to decide which clothes to make.
How should they frame their problem in ML terms?

Ideal outcome:
Model's goal:
Model output:
Success metrics:

A fashion firm wants to sell more clothes. Someone suggests using ML to determine which clothes the firm should manufacture. They think they can train a model to determine which type of clothes are in fashion. After they train the model, they want to apply it to their catalog to decide which clothes to make.

How should they frame their problem in ML terms?

Ideal outcome: Determine which products to manufacture.

Model's goal: Predict which articles of clothing are in fashion.

Model output: Binary classification, in_fashion, not_in_fashion

Success metrics: Sell seventy percent or more of the clothes made.

If you had to prioritize improving one of the areas below in your machine learning project, which would have the most impact?

A. A deeper network
B. A more clever loss function
C. The quality and size of your data
D. Using the latest optimization algorithm

Take a guess:

In your machine learning project, how much time will you typically spend on data preparation and transformation?

A. Less than half of the project time
B. More than half of the project time

Data Management

Data Management

INPUT

Set of Requirements that model should satisfy.

OUTPUT

1. Training Dataset
2. Testing Dataset

Data Mangement

ACTIVITIES

1. Data Collection
2. PreProcessing
3. Augmentation
4. Analysis

PROPERTIES

1. Relevant
2. Complete
3. Balanced
4. Accurate

Model Learning

What is a "model" in machine learning?

A. A model is a piece of computer hardware
B. A model is a mathematical relationship derived from data that an ML system uses to make predictions
C. A model is a smaller representation of the thing you're studying.

Model Learning

INPUT

Training dataset

OUTPUT

1. Machine Learned model
2. Performance deficit report used to inform any remedial data management activities.

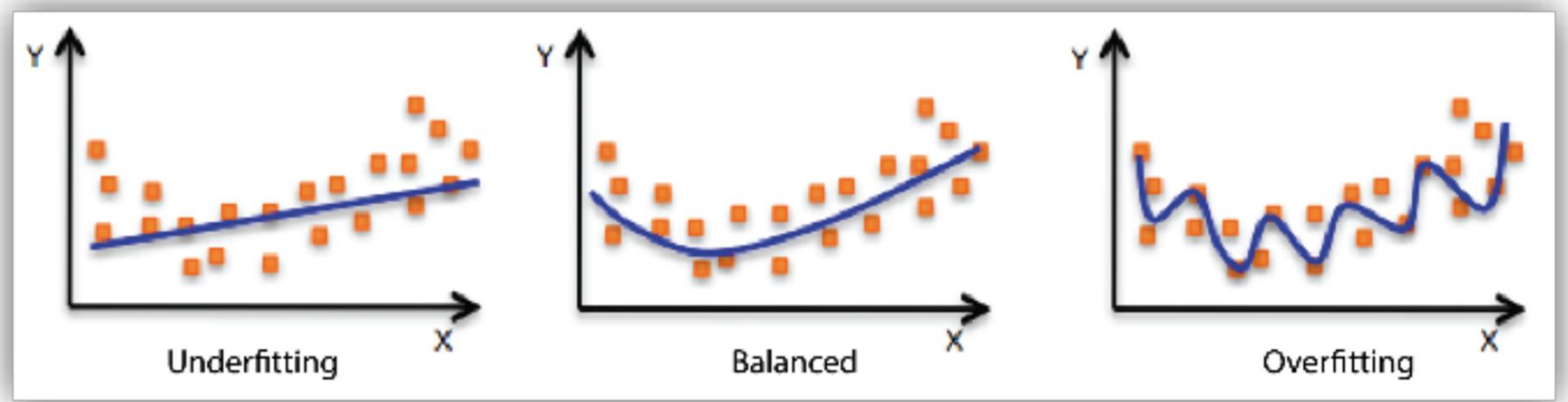
Model Learning

ACTIVITIES

1. Model Selection
2. Training
3. Hyperparameter selection
4. Transfer Learning

PROPERTIES OF MODEL

1. Performant
2. Robust
3. Reusable
4. Interpretable


Y
X
Underfitting
Y
X
Balanced
Y
X
Overfitting

Model Verification

Model Verification/Evaluation

INPUT

Trained Model

OUTPUT

Verified Model

Verification Result that provides information to determine if the model is suitable for intended applications.

Model Verification

ACTIVITIES

1. Requirement Encoding
2. Test-based Verification
3. Formal Verification

DESIRED PROPERTIES OF VERIFICATION RESULTS

1. Comprehensive
2. Contextually Relevant
3. Comprehensible

Model Deployment

Model Deployment

INPUT

Verified Model

Verification evidence

OUTPUT

Model suitably deployed within a system

Model Deployment

PROPERTIES

1. Fit -for-purpose
2. Tolerable
3. Adaptable

Monitoring and Maintainence

- After deploying the model to production we need to constantly monitor and improve the system.
- We will be monitoring model metrics, hardware and software performance, and customer satisfaction.
- Notification about the anomalies, reduced model and system performance, and bad customer reviews.
- After we get a reduced performance alert, we will assess the issues and try to train the model on new data or make changes to model architectures.

Terminologies

Algorithm

a set of rules that a machine follows to achieve a particular goal.

An algorithm can be considered as a recipe that defines the inputs, the output and all the steps needed to get from the inputs to the output.

Cooking recipes are algorithms where the ingredients are the inputs, the cooked food is the output, and the preparation and cooking steps are the algorithm instructions.

Learner or **Machine Learning Algorithm**

the program used to learn a machine learning model from data.

Machine Learning Model

the learned program that maps inputs to predictions.

Dataset

a table with the data from which the machine learns. The dataset contains the features and the target to predict.

Instance

a row in the dataset.

Features

the inputs used for prediction or classification. A feature is a column in the dataset.

Target

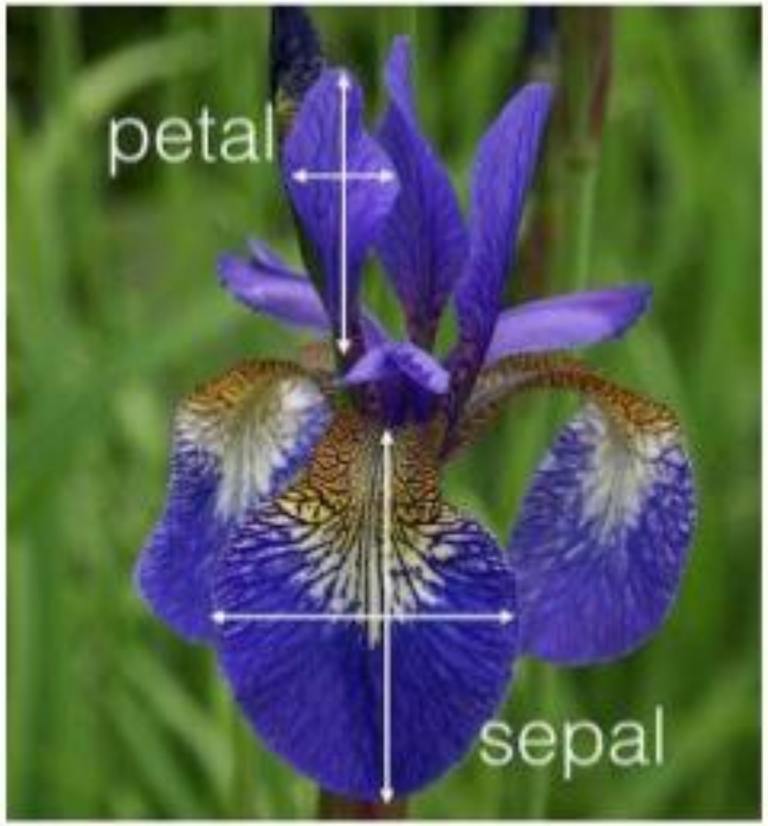
the information the machine learns to predict.

Machine Learning Task

the combination of a dataset with features and a target. Depending on the type of the target, the task can be for example classification, regression, survival analysis, clustering, or outlier detection.

Prediction

what the machine learning model “guesses” what the target value should be based on the given features.



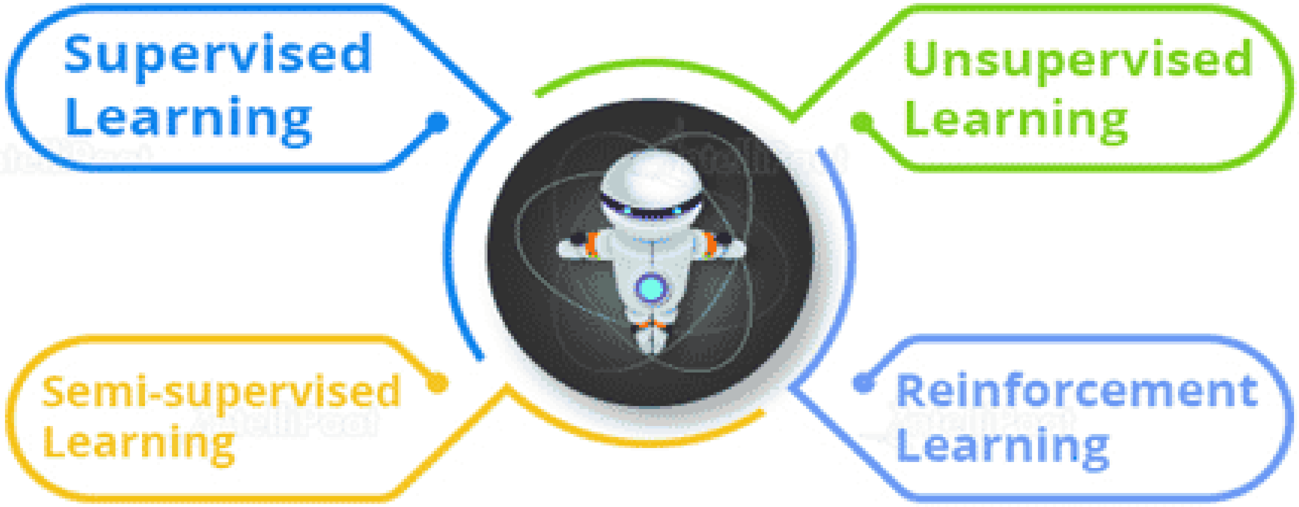
Training / test data

Features | Labels

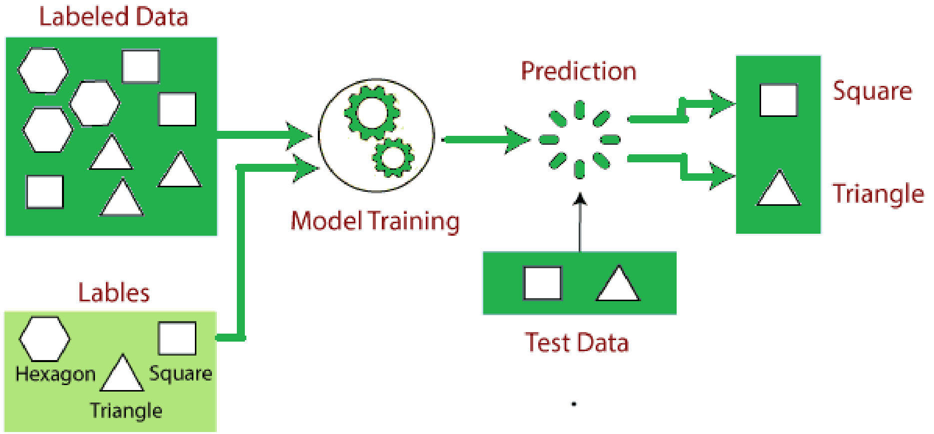
Sepal length	Sepal width	Petal length	Petal width	Species
5.1	3.5	1.4	0.2	Iris setosa
4.9	3.0	1.4	0.2	Iris setosa
7.0	3.2	4.7	1.4	Iris versicolor
6.4	3.2	4.5	1.5	Iris versicolor
6.3	3.3	6.0	2.5	Iris virginica
5.8	3.3	6.0	2.5	Iris virginica

ML Types

ML Types

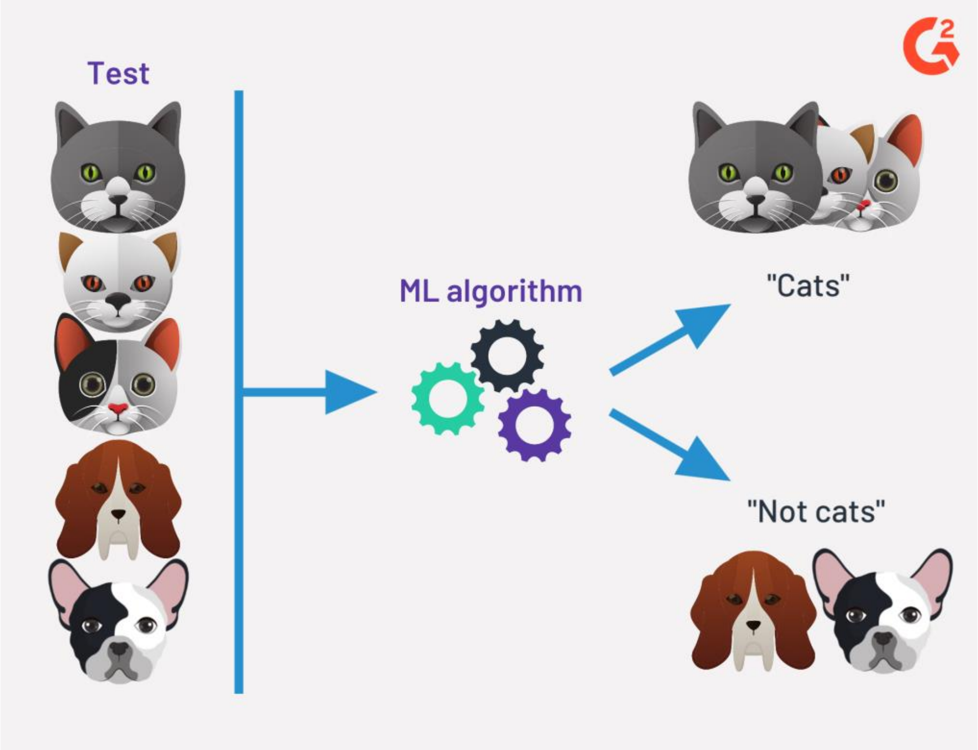

Supervised Learning
Unsupervised Learning
Semi-supervised Learning
Reinforcement Learning

Supervised Learning



- Image Segmentation
- Medical Diagnosis
- Fraud Detection
- Spam detection
- Speech Recognition

Unsupervised Learning



- ❑ Network Analysis
- ❑ Recommendation Systems
- ❑ Anomaly Detection

Inclusys Org Foundation
Empowering the Neurodivergent

Semi-supervised Learning

Data

X1	X2	Label
1	2	1
0	0.3	0
3	4	0
1.4	2.3	1
4.3	9.2	?
4.1	2	?
4.3	9.8	?
9.8	21	?
3.2	9	?
5	2	?
4.3	8.7	?
3.2	8.7	?

Partially Labelled

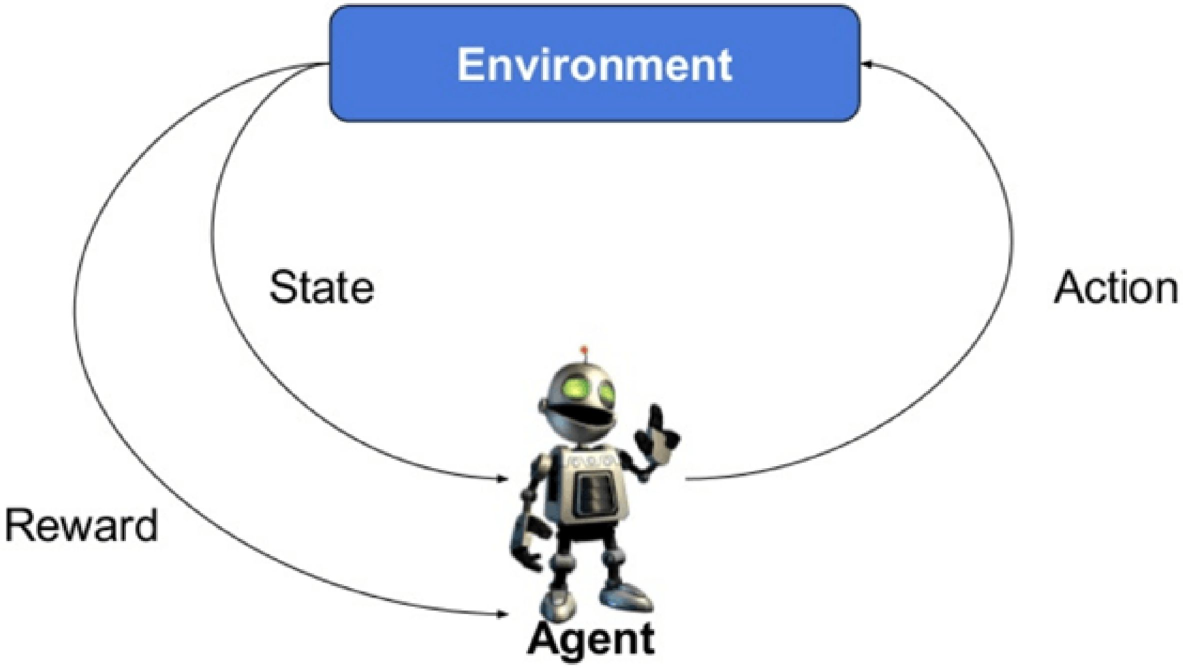
Mostly Unlabeled

Supervised Learning Model

Unsupervised Learning Model

Semi-Supervised Learning Model

Reinforcement Learning

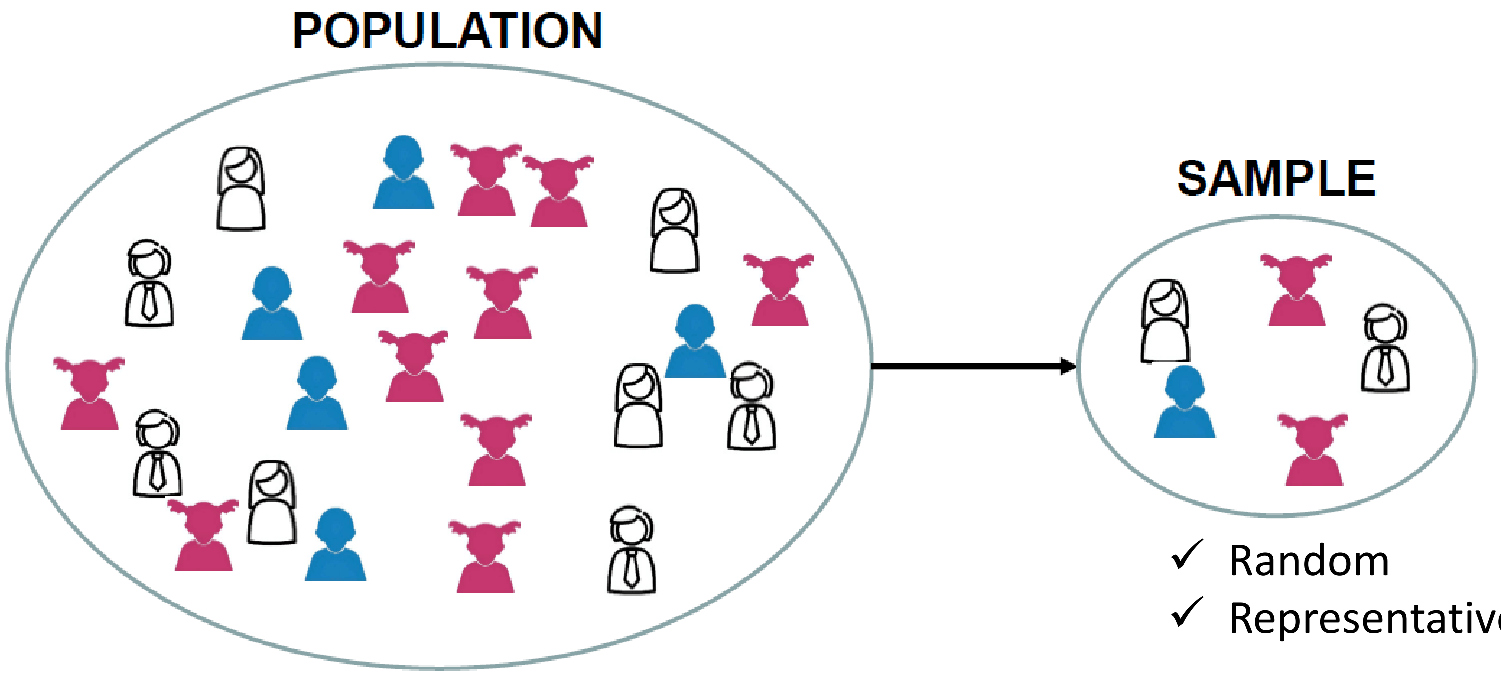


- ❑ Video Games
- ❑ Robotics
- ❑ Text Mining

Data Preprocessing

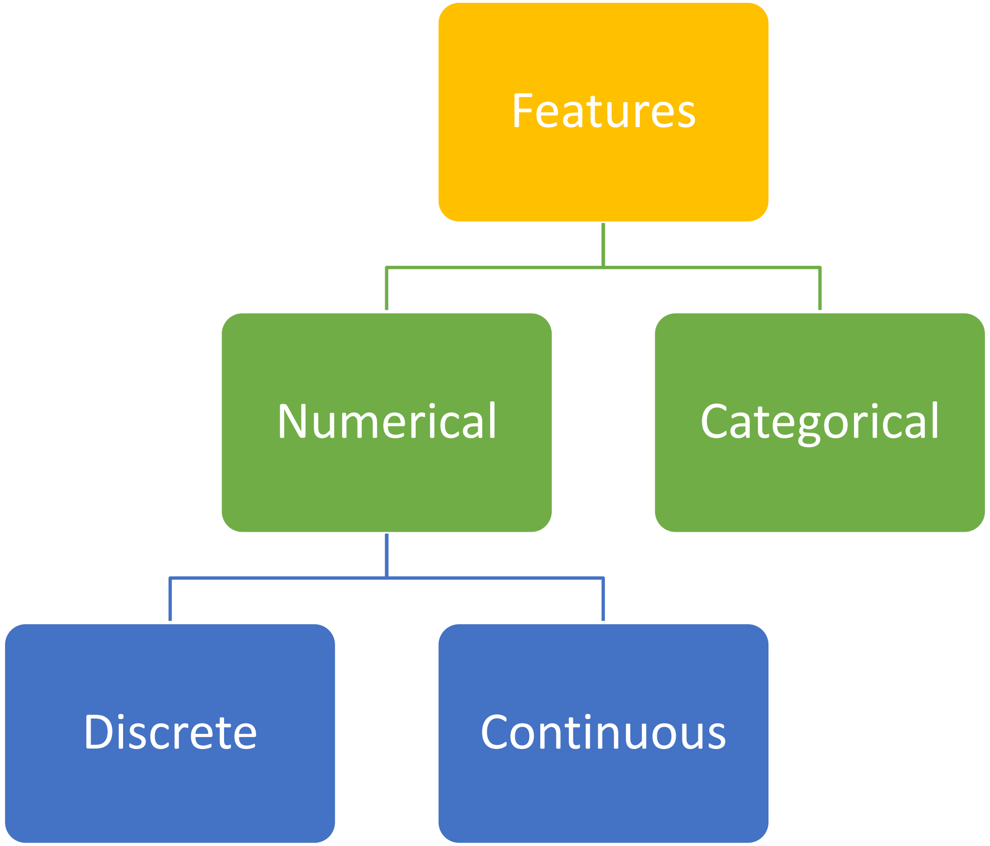
Exploratory Data Analysis

Population and Sample

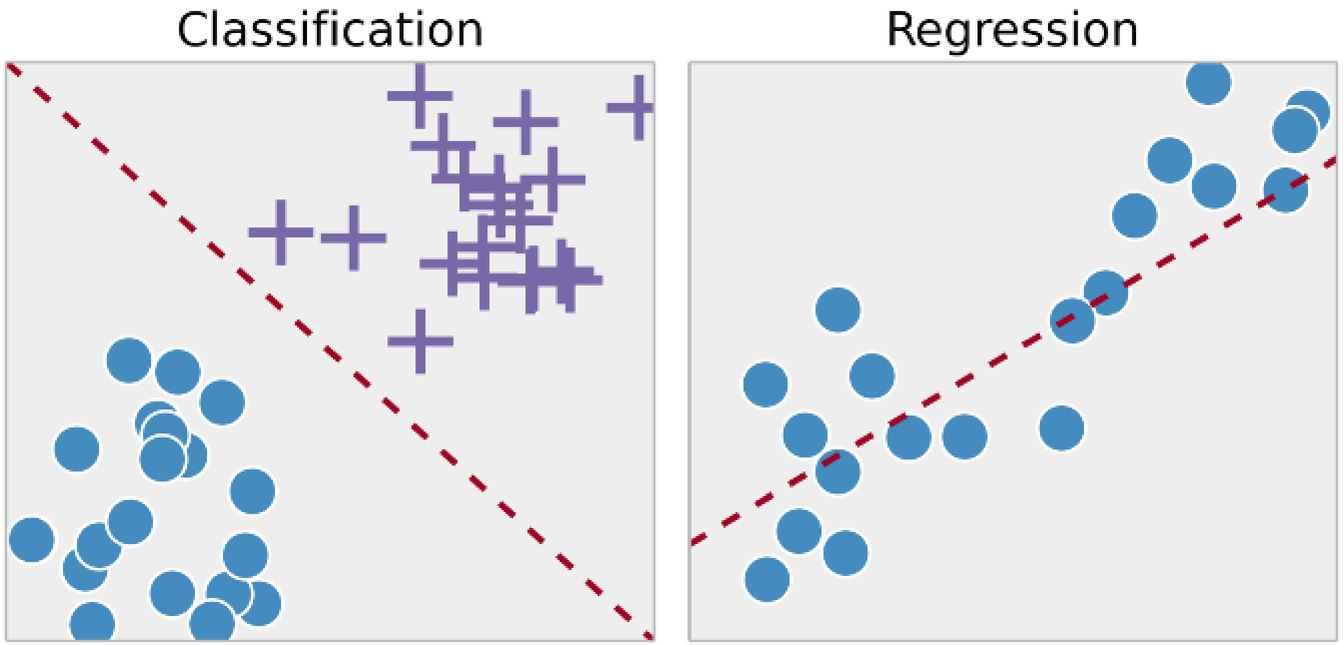


- ✓ Random
- ✓ Representative

Types of features : Data type



Types of Supervised Machine Learning



Determining whether or not someone will be a defaulter of the loan

predict the house price from training data

Types of Features : Measurement Levels

QUALITATIVE

1. Nominal
2. Ordinal

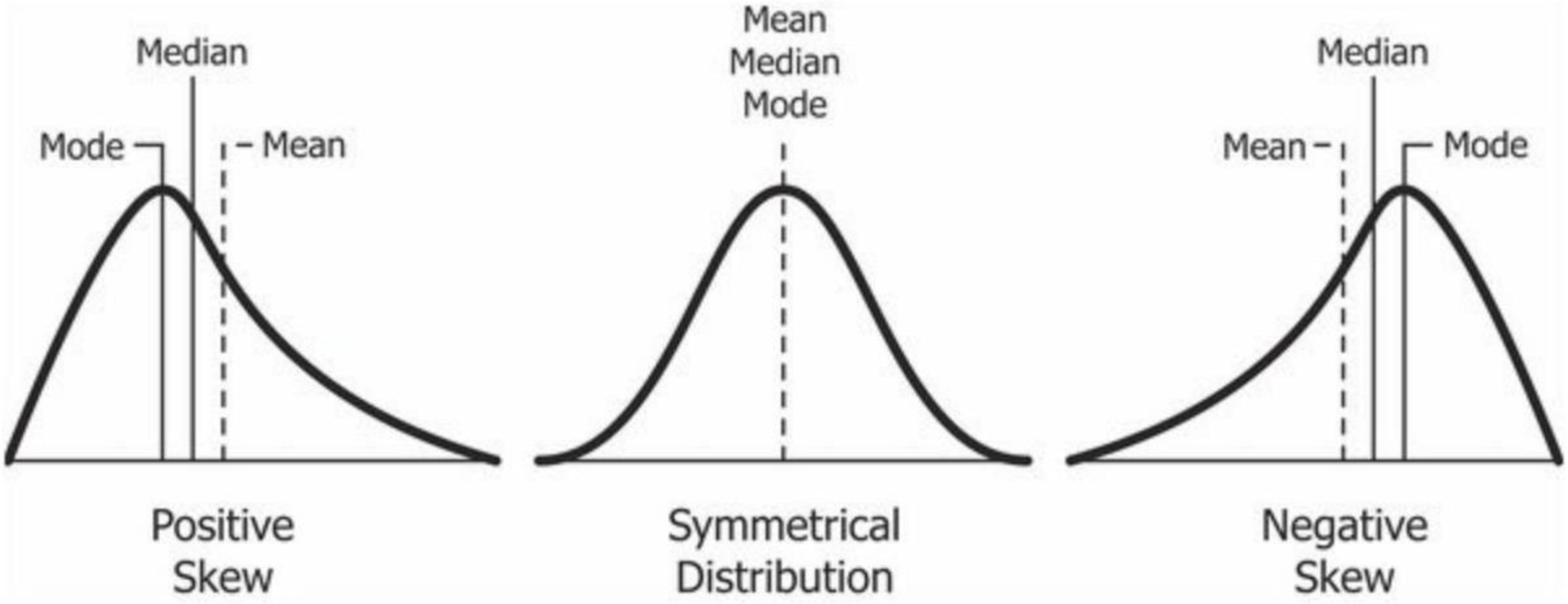
QUANTITATIVE

1. Interval
2. Ratio

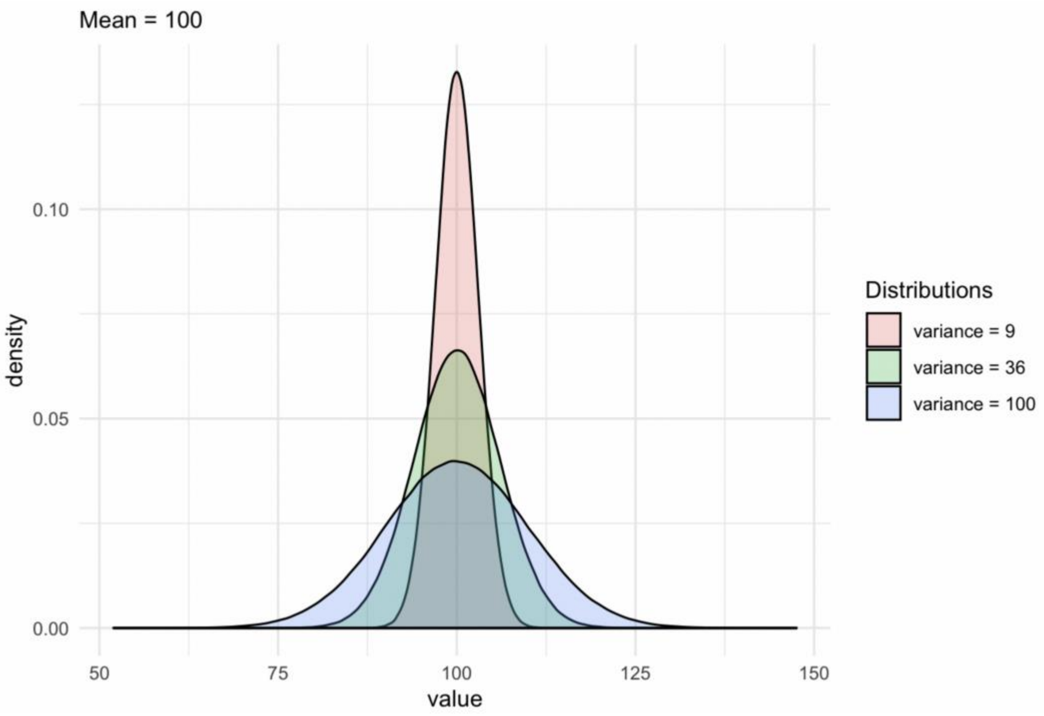
Measures of Central Tendency

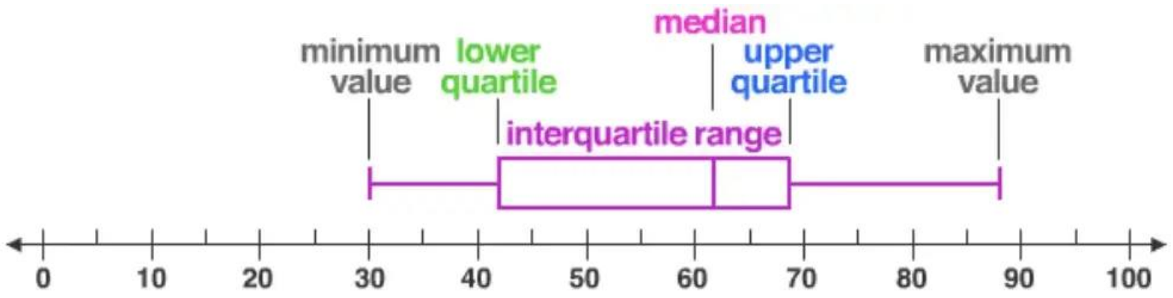
1. Mean
2. Median
3. Mode

Measures of Assymmetry



Measures of Variability





Share your feedback

https://forms.gle/8nX9pX9Aoxtkb1Ap7

Thank You!